
**SYSTEM AND METHOD FOR DIVIDING DATA INTO
PREDOMINANTLY FIXED-SIZED CHUNKS SO THAT
DUPLICATE DATA CHUNKS MAY BE IDENTIFIED**

ABSTRACT OF THE INVENTION

A data chunking system divides data into predominantly fixed-sized chunks such that duplicate data may be identified. The data chunking system may be used to reduce the data storage and save network bandwidth by allowing storage or transmission of primarily unique data chunks. The system may also be used to increase reliability in data storage and network transmission, by allowing an error affecting a data chunk to be repaired with an identified duplicate chunk. The data chunking system chunks data by selecting a chunk of fixed size, then moving a window along the data until a match to existing data is found. As the window moves across the data, unique chunks predominantly of fixed size are formed in the data passed over. Several embodiments provide alternate methods of determining whether a selected chunk matches existing data and methods by which the window is moved through the data. To locate duplicate data, the data chunking system remembers data by computing a mathematical function of a data chunk and inserting the computed value into a hash table.